



## CASE STUDY

# DATA DIGITIZATION USING MS AZURE OCR-BASED SMARTCAPTURE FOR A DIAGNOSTICS FIRM

Enabled scientists with access to 99.5% accurate digitized documents of manually processed records

## Client Background

Our client is a leading global company that manufactures a wide array of innovative medical diagnostic assays. The company had about 50,000 Batch Record documents in scanned format from 8 product families. Each product family had 2 types of documents i.e. PBR (Production Batch Record) and FLR (Filling and Labelling Record). The client wanted a solution to digitize and index this information to equip scientists with real-time access to critical data records for quick statistical analysis. The key objectives included:

- Make the above documents searchable
- Extract defined entities from scanned Batch Record documents, including enhancements for managing handwritten entities
- Build UI proofreading and fast retrieval for analysis
- Ensure extraction accuracy within the desired range

## Xoriant Solution | Key Contributions

Xoriant used its very own state-of-the-art intuitive document digitization solution SmartCapture developed using MS Azure OCR. The solution made reports searchable for clinical scientists with limited IT knowledge. We performed data classification, data extraction, and the necessary quality checks to convert the manual records into accurate digital records.

**Handled document classification, data extraction using Xoriant SmartCapture** by studying a set of client's documents from diverse families. A multistep machine learning-based classification identified each document's type, version, page number, and document structure for data extraction of printed text by

## KEY BENEFITS

- Delivered **99.5%** accuracy levels in digitized documents with Data Quality Improvement efforts
- Enabled scientists with quick search of historical documents, data visualization to make product improvements using analytics
- Minimized information search time from hours to seconds and improved investigation efficiency with a centralized data storage and retrieval system
- Reduced the time needed for data mining and allowed more time (scientists) to work on investigations with digital transformation

applying OCR techniques. Each document was digitized using specific configurations.

**Performed quality checks for data accuracy using SmartCapture solution** for predicting the attributes that required corrections. The outliers were checked and rectified manually to ensure maximum data accuracy in the digitized documents.

**Developed UI for quick document access to scientists** with the ability to access the rendered form of the actual data in a familiar format. The extracted data can be exported by the scientists in an Excel or PDF format for further analysis.

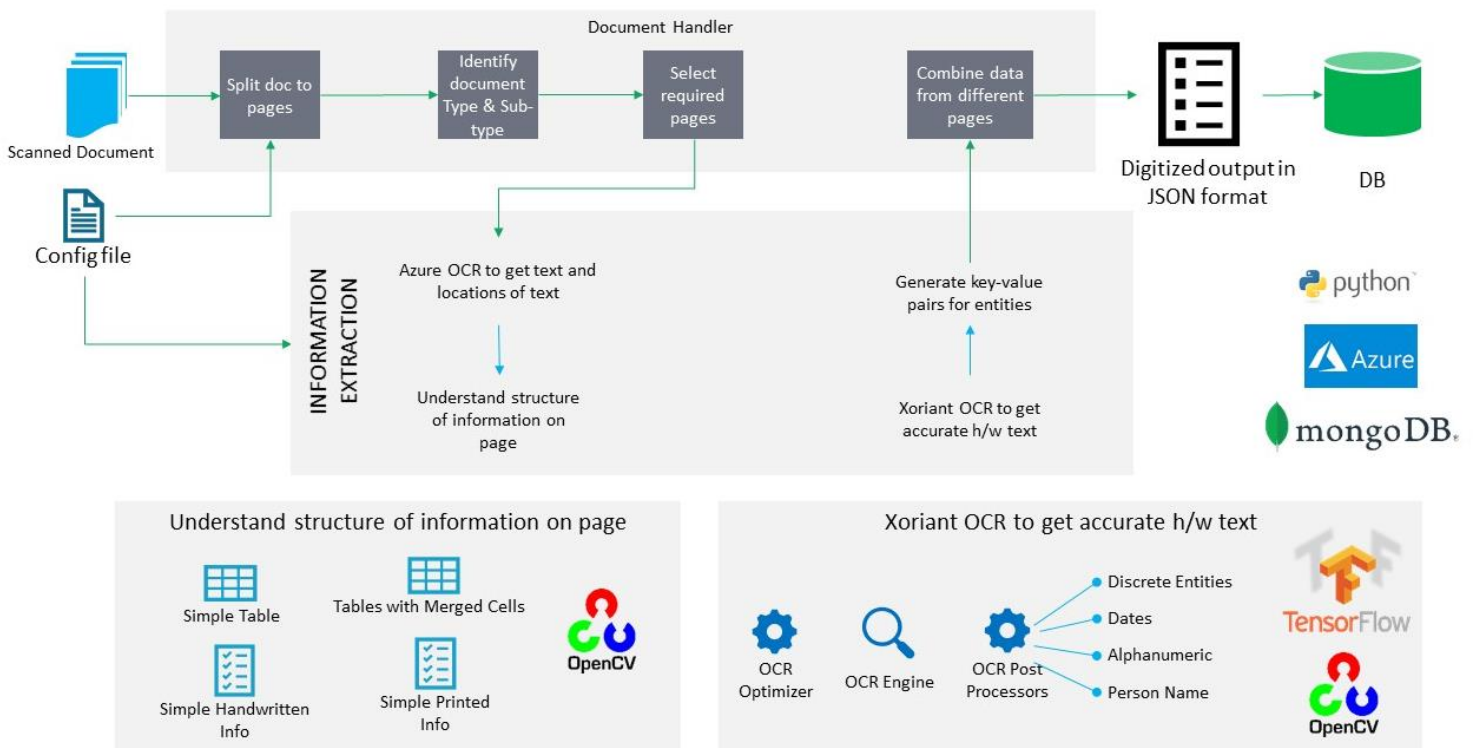
## Client Testimonial



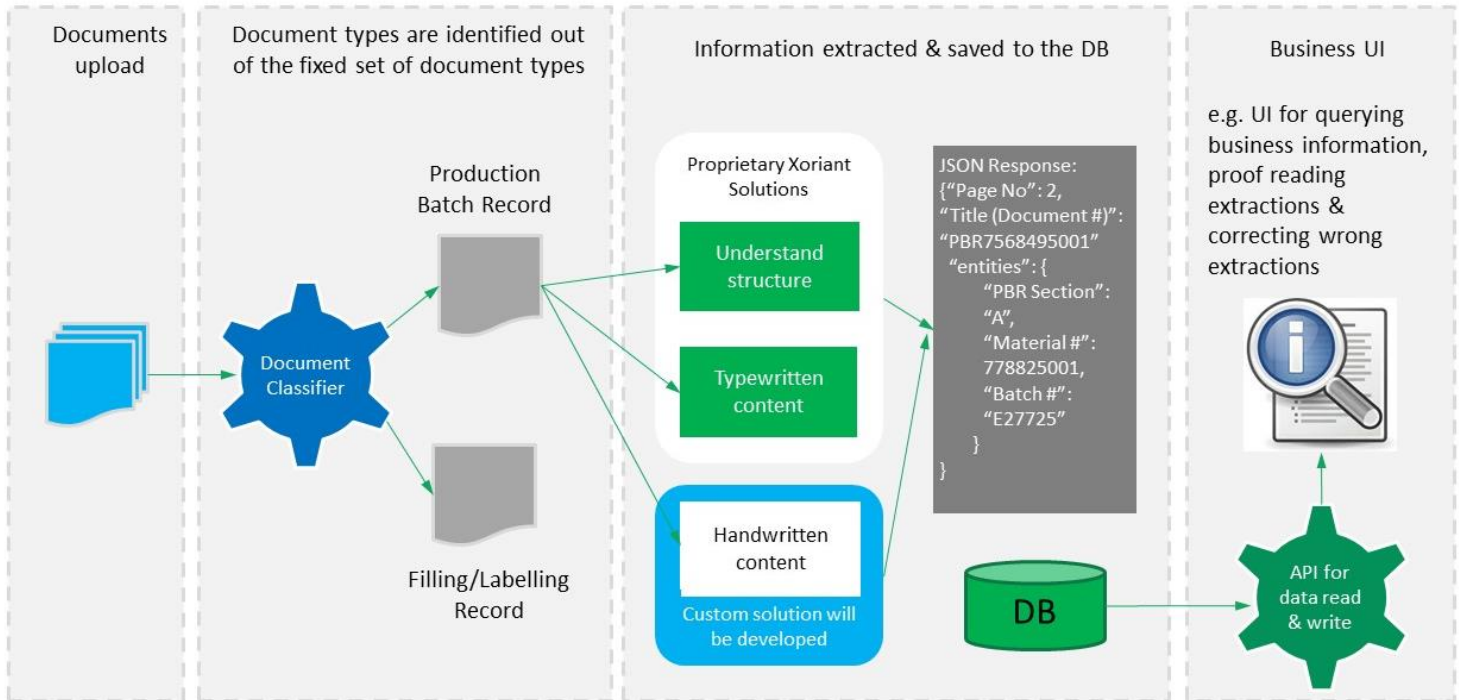
*Xoriant's SmartCapture enabled our team to access, analyze millions of data entries by resolving the complexities in this challenging project. Instead of spending hours sifting through thousands of pages, our scientists can now find specific information in seconds. Xoriant's remarkable tech expertise helped our scientists to use their saved time to complete investigation and extend analysis.*



## High Level Architecture



### XORIENT SMARTCAPTURE SOLUTION



### Technology Stack

**Xoriant SmartCapture | AWS (S3, Redshift, EC2) | SFDC | Oracle Fusion | Tableau  
 CloudIO (FLOW) | Python | Computer Vision (OpenCV) | Object Detection (inception model)  
 Image Processing (OpenCV) | Tesseract OCR | Azure OCR | Handwritten Text Detection Models  
 No-SQL DB (Elasticsearch/MongoDB) | Virtual Machines on Azure | Kubernetes and Containers**



Xoriant is a product engineering, software development and technology services company, serving technology startups as well as mid-size to large corporations. We offer a flexible blend of onsite, offsite and offshore services from our eight global delivery centers with over 4000 software professionals. Xoriant has deep client relationships spanning over 30 years with various clients ranging from startups to Fortune 100 companies.